

Research Article

Modelling Nonstationary Gene Regulatory Processes

Marco Grzegorzczk,¹ Dirk Husmeier,² and Jörg Rahnenführer¹

¹Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany

²Biomathematics and Statistics Scotland, JCMB, The King's Buildings, Edinburgh EH93JZ, UK

Correspondence should be addressed to Marco Grzegorzczk, grzegorzczk@statistik.tu-dortmund.de

Received 18 December 2009; Accepted 29 April 2010

Academic Editor: Yves Van de Peer

Copyright © 2010 Marco Grzegorzczk et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An important objective in systems biology is to infer gene regulatory networks from postgenomic data, and dynamic Bayesian networks have been widely applied as a popular tool to this end. The standard approach for nondiscretised data is restricted to a linear model and a homogeneous Markov chain. Recently, various generalisations based on changepoint processes and free allocation mixture models have been proposed. The former aim to relax the homogeneity assumption, whereas the latter are more flexible and, in principle, more adequate for modelling nonlinear processes. In our paper, we compare both paradigms and discuss theoretical shortcomings of the latter approach. We show that a model based on the changepoint process yields systematically better results than the free allocation model when inferring nonstationary gene regulatory processes from simulated gene expression time series. We further cross-compare the performance of both models on three biological systems: macrophages challenged with viral infection, circadian regulation in *Arabidopsis thaliana*, and morphogenesis in *Drosophila melanogaster*.

1. Introduction

An important objective in systems biology is to infer regulatory networks from postgenomic data. Bayesian networks have been widely applied as a popular tool to this end—see, for example, [1]—and novel fast Markov chain Monte Carlo (MCMC) algorithms can be applied to systematically search the space of network structures for those that are most consistent with the data [2, 3]. One reason for the popularity of Bayesian networks has been the tractability of the marginal likelihood of the network structure. This term describes how well the model structure explains the data. Its computation is usually challenging, as it requires the solution of an integral over the entire parameter space. To obtain a closed-form expression, two probabilistic models with their respective conjugate prior distributions have been employed: the multinomial distribution with the Dirichlet prior (BDe) and the Gaussian distribution with the normal-Wishart prior (BGe). These approaches are restricted in that they either require a data discretisation (BDe: Bayesian Dirichlet equivalence score) or can only capture linear relationships (BGe: Bayesian Gaussian equivalence score). Recently, a

generalisation of the BGe model based on a combination of a mixture model and the allocation sampler has been proposed [4], the Bayesian Gaussian Mixture (BGM) model. In the BGM model, data points are assigned to different compartments (subsets of the data) with the allocation sampler [5]. Model parameters (and their distributions) are allowed to differ between compartments, while information is shared among the compartments via a common network structure. Given the network structure each compartment is modelled separately and independently with the Gaussian BGe model.

The present work proposes a modification of the BGM model, which is more suitable for dynamic gene regulatory networks and gene expression time series. The proposed model, which we refer to as the BGM_D model, replaces the free allocation model by a multiple changepoint process, as, for example, employed in [6], to divide the time series into segments and thereby take the time structure into account.

In a preliminary study [7], we focused on synthetic data from small network domains. We found that the new model avoids spurious self-feedback loops if feedback loops, such as $X(t) \rightarrow X(t+1)$ (e.g., in molecular biology: transcription

factors regulating their own transcription), are not excluded from the analysis. In this paper, we present the mathematical details of the new BGM_D model in greater depth, and we demonstrate that BGM_D yields better inference results for real biological systems. To this end, we analyse two gene expression time series from macrophages and *Arabidopsis thaliana* and cross-compare the inference results of the BGM model and the new changepoint process BGM_D model.

Furthermore, we combine both models (the free allocation model and the changepoint process) with discrete Bayesian network methodology and we compare the performance of both models on inferring the morphogenetical stages of muscle development in *Drosophila melanogaster* from binary gene expression time series.

We note that our modelling paradigm is complementary to other recently proposed approaches. A nonlinear nondiscretised model based on heteroscedastic regression has been proposed in [8]. However, this approach no longer allows the marginal likelihood to be obtained in closed-form and requires a restrictive approximation (the Laplace approximation, that is, an approximation based on a 2nd order Taylor series approximation of the log likelihood) to be adopted. Another nonlinear model based on node-specific Gaussian mixture models has been proposed in [9, 10]. Again, the marginal likelihood is intractable. The authors resort to the Bayesian information criterion (BIC) of [11] for model selection, which is only a good approximation to the marginal likelihood in the limit of very large data sets. A more flexible approach based on changing graphs between changepoints has been proposed in [12, 13]. Conceptually, the assumption of changing networks is reasonable for some biological scenarios, like morphogenesis, where gene-regulatory processes have been measured over a long period of time. However, for cellular processes on a shorter time scale, it is questionable whether it is the network structure rather than just the strength of the regulatory interactions that changes with time. For example, it is not the principle ability of a transcription factor to potentially bind to the promoter of a gene and thereby initiate transcription (i.e., not the network structure), but the extent to which this happens (i.e., the regulatory interaction strength). We therefore argue that, especially for short time series, it is more appropriate to leave the network structure invariant among segments and to allow the interaction parameters to vary with time by modelling the time segments (between changepoints) separately and independently with the Gaussian BGe model. The systematic sharing of information among segments via a common network structure also reduces potential problems with model overflexibility and overfitting, which a more flexible approach that allows for different network structures is susceptible to when the time series are short.

2. Methods

2.1. The Dynamical BGe Network. Dynamical Bayesian networks (DBNs) are flexible models for representing probabilistic relationships among interacting variables X_1, \dots, X_N . The graph \mathcal{G} of a DBN describes the relationships between the

variables (nodes), which have been measured at equidistant time points $t = 1, \dots, m$, in the form of conditional probability distributions. An edge pointing from X_i to X_j means that the realisation of X_j at time point t , symbolically: $X_j(t)$, is influenced by the realisation of X_i at time point $t - 1$, symbolically: $X_i(t - 1)$. Effectively there is a bipartite graph structure between two time steps t and $t + 1$ so that the acyclicity constraint—known from static Bayesian networks—is guaranteed to be satisfied. Therefore, in principle, each node can be its own parent node in DBNs. Such self-feedback loops $X_n(t - 1) \rightarrow X_n(t)$ model autocorrelations and it depends on the application whether such self-feedback loops should be valid edges or ruled out altogether to focus the inference on the interactions between genes. $\pi_n = \pi_n(\mathcal{G})$ denotes the parent (node) set of node X_n in \mathcal{G} ; that is, the set of all nodes from which an edge points to node X_n in \mathcal{G} . Given a data set \mathbf{D} , where $\mathbf{D}_{n,t}$ and $\mathbf{D}_{(\pi_n,t)}$ are the t th realisations $X_n(t)$ and $\pi_n(t)$ of X_n and π_n , respectively, DBNs are based on the following homogeneous Markov chain expansion:

$$P(\mathbf{D} | \mathcal{G}, \boldsymbol{\theta}) = \prod_{n=1}^N \prod_{t=2}^m P(X_n(t) = \mathbf{D}_{n,t} | \pi_n(t-1) = \mathbf{D}_{(\pi_n,t-1)}, \boldsymbol{\theta}_n), \quad (1)$$

where $\boldsymbol{\theta}$ is the total parameter vector, composed of subvectors $\boldsymbol{\theta}_n$. $\boldsymbol{\theta}_n$ specifies the n th local conditional distributions $P(X_n(t) | \pi_n(t-1), \boldsymbol{\theta}_n)$ in the factorisation. The BGe model specifies the distributional form $P(\mathbf{D} | \mathcal{G}, \boldsymbol{\theta})$ as multivariate Gaussian distribution, and it assumes a normal-Wishart distribution as prior distribution $P(\boldsymbol{\theta} | \mathcal{G})$ [14]. The local probability distributions $P(X_n(t) | \pi_n(t-1), \boldsymbol{\theta}_n)$ are then given by conditional linear Gaussian distributions. For discrete data \mathbf{D} the BDe model specifies the distributional form of the likelihood $P(\mathbf{D} | \mathcal{G}, \boldsymbol{\theta})$ as a set of multinomial distributions, and assumes a Dirichlet distribution as the prior $P(\boldsymbol{\theta} | \mathcal{G})$ for the unknown parameters [15]. In principle, the BDe model yields a higher modelling flexibility than the BGe model, but BDe requires a data discretisation that usually incurs a substantial information loss. Under fairly weak conditions imposed on the parameter vector (prior independence and modularity) and the prior distribution $P(\boldsymbol{\theta})$ (conjugacy), so that $P(\boldsymbol{\theta} | \mathcal{G}) = \prod_{n=1}^N P(\boldsymbol{\theta}_n | \pi_n)$, where $\pi_n = \pi_n(\mathcal{G})$, the parameters can be integrated out analytically. For both scoring metrics BGe and BDe [16], the marginal likelihood then satisfies the same expansion rule as the Bayesian network with fixed parameters:

$$P(\mathbf{D} | \mathcal{G}) = \int P(\mathbf{D} | \mathcal{G}, \boldsymbol{\theta}) P(\boldsymbol{\theta} | \mathcal{G}) d\boldsymbol{\theta} = \prod_{n=1}^N \Psi(\mathbf{D}_n^{\pi_n}), \quad (2)$$

where

$$\Psi(\mathbf{D}_n^{\pi_n}) = \int \prod_{t=2}^m P(X_n(t) = \mathbf{D}_{n,t} | \pi_n(t-1) = \mathbf{D}_{(\pi_n,t-1)}, \boldsymbol{\theta}_n) \times P(\boldsymbol{\theta}_n | \pi_n) d\boldsymbol{\theta}_n, \quad (3)$$

and $\mathbf{D}_n^{\pi_n} := \{(\mathbf{D}_{n,t}, \mathbf{D}_{\pi_n,t-1}) : 2 \leq t \leq m\}$ denotes the subset of the data pertaining to node X_n and its parent set π_n . For the Gaussian BGe model the (local) factors $\Psi(\mathbf{D}_n^{\pi_n})$ in (3) can be computed in closed-form, according to (15) and (24) in Geiger and Heckerman [14]. For the discrete multinomial BDe model, the (local) factors can be computed in closed form according to (19) in [16].

2.2. Gaussian Mixture Bayesian Network Model. In the Gaussian BGe model, the local distributions $P(X_n(t) | \pi_n(t-1), \boldsymbol{\theta}_n)$ are conditional linear Gaussian distributions. We generalise the BGe model by the introduction of a latent allocation vector \mathbf{V} , which assigns the data points to K different mixture components, where K is inferred from the data by applying changepoint birth and death moves, along the lines of the changepoint model (e.g., see [6, 17]). As in the BGM model [4], conditional on the latent vector \mathbf{V} , a separate BGe score can be computed for each of the K mixture components.

The allocation vector \mathbf{V} of size $m - 1$ describes the allocation of the time points $t = 2, \dots, m$ to the K components. $\mathbf{D}^{(\mathbf{V},k)}$ denotes all time points that are allocated to component k . The posterior probability of \mathcal{G} , \mathbf{V} , and K is proportional to the joint distribution

$$P(\mathcal{G}, \mathbf{V}, K | \mathbf{D}) = \frac{P(\mathcal{G}, \mathbf{V}, K, \mathbf{D})}{P(\mathbf{D})} \propto P(\mathcal{G}, \mathbf{V}, K, \mathbf{D}), \quad (4)$$

and the joint distribution can be factorised as follows:

$$P(\mathcal{G}, \mathbf{V}, K, \mathbf{D}) = P(K)P(\mathbf{V} | K)P(\mathcal{G})P(\mathbf{D} | \mathcal{G}, \mathbf{V}, K), \quad (5)$$

where

$$P(\mathbf{D} | \mathcal{G}, \mathbf{V}, K) = \prod_{k=1}^K P(\mathbf{D}^{(\mathbf{V},k)} | \mathcal{G}) = \prod_{k=1}^K \prod_{n=1}^N \Psi(\mathbf{D}_n^{(\mathbf{V},k), \pi_n}), \quad (6)$$

and $\mathbf{D}_n^{(\mathbf{V},k), \pi_n} := \{(\mathbf{D}_{n,t}, \mathbf{D}_{\pi_n,t-1}) | t \in \{2, \dots, m\} : \mathbf{V}(t) = k\}$ denotes the realisations of node X_n and its parent set π_n for those time points that have been allocated to the k th component. It can be seen from these equations that \mathbf{V} acts as a filter which divides the data \mathbf{D} into K different compartments $\mathbf{D}^{(\mathbf{V},k)}$ ($k = 1, \dots, K$), for which separate independent BGe scores can be computed in closed-form using (2) and (3). The BGM counterpart of (3) is given by:

$$\begin{aligned} & \Psi(\mathbf{D}_n^{(\mathbf{V},k), \pi_n}) \\ &= \int \prod_{t: \mathbf{V}(t)=k} P(X_n(t) = \mathbf{D}_{n,t} | \pi_n(t-1) = \mathbf{D}_{\pi_n,t-1}, \boldsymbol{\theta}_n) \\ & \quad \times P(\boldsymbol{\theta}_n | \pi_n) d\boldsymbol{\theta}_n. \end{aligned} \quad (7)$$

For instance, if we have $m = 11$ time points and one changepoint between t_6 and t_7 so that \mathbf{V} assigns the time points t_2, \dots, t_6 to the first and the remaining time points t_7, \dots, t_{11} to the second segment, then separate local BGe scores are computed for the data subsets $\mathbf{D}_n^{(\mathbf{V},1), \pi_n} := \{(\mathbf{D}_{n,t}, \mathbf{D}_{\pi_n,t-1}) | 2 \leq t \leq 6\}$ and $\mathbf{D}_n^{(\mathbf{V},2), \pi_n} :=$

$\{(\mathbf{D}_{n,t}, \mathbf{D}_{\pi_n,t-1}) | 7 \leq t \leq 11\}$, according to (15) and (24) in Geiger and Heckerman [14].

When a data compartment $\mathbf{D}^{(\mathbf{V},k)}$ is empty, then we set the factors $\Psi(\mathbf{D}_n^{(\mathbf{V},k), \pi_n})$ equal to 1 ($n = 1, \dots, N$). For $P(\mathcal{G})$, we take a uniform distribution over all graph structures subject to a fan-in restriction of $|\pi_n| \leq \mathcal{F}$. For $P(K)$, we take a truncated Poisson distribution with $\lambda = 1$ restricted to $1 \leq K \leq K^*$ as prior. We note that the MCMC inference scheme, which we will discuss in Section 2.3, does not sample \mathbf{V} directly, but is based on local modifications of \mathbf{V} based on changepoint birth, death, and reallocation moves. That is, different from the free allocation in the BGM model [4], we change the assignment of data points to components via a changepoint process (e.g., see [6, 17]). This reduces the complexity of the allocation space substantially and incorporates our prior knowledge that adjacent time points are likely to be assigned to the same component. We will refer to the new changepoint BGM model as the BGM_D model.

We identify K with $K - 1$ changepoints: b_1, \dots, b_{K-1} on the continuous interval $[2, m]$, and for notational convenience we introduce the pseudo-changepoints $b_0 = 2$ and $b_K = m$. The observation at time point t is assigned to the k th component, symbolically $\mathbf{V}(t) = k$, if $b_{k-1} \leq t < b_k$. Following Green [17], we assume that the changepoints are distributed as the even-numbered order statistics of $L := 2(K - 1) + 1$ points u_1, \dots, u_L uniformly and independently distributed on the interval $[2, m]$. The motivation for this prior, instead of taking $K - 1$ uniformly distributed points, is to encourage *a priori* an equal spacing between the changepoints, that is, to discourage mixture components that contain only a short compartment of the time series. We note that the even-numbered order statistics prior on the changepoint locations induces a prior distribution on the allocation vector \mathbf{V} . Deriving a closed-form expression is involved but not required, as our MCMC inference scheme does not sample \mathbf{V} directly. The inference is based on local modifications of \mathbf{V} . In particular, we employ three different move types: (i) an additional component ($K \rightarrow K + 1$) can be generated by setting a new changepoint on the interval $[2, m]$ (changepoint birth move), (ii) the number of components can be decremented ($K \rightarrow K - 1$) by removing one of the existing changepoints (changepoint death move), and finally, (iii) the allocation vector \mathbf{V} can be changed without affecting the number of components K by changing the position of one of the existing changepoints (changepoint reallocation move). For the acceptance probabilities of these local moves, only $P(\mathbf{V} | K)$ ratios, which are straightforward to compute, are required.

2.3. Multinomial Mixture Bayesian Network Model. The discrete multinomial BDe model can be generalised in analogy to the continuous Gaussian Bge model. As before, the allocation vector \mathbf{V} divides the data into K compartments, and each compartment of the data is modelled separately and independently with the multinomial BDe scoring metric. We note that closed-form solutions for the $\Psi(\cdot)$ -terms in (7) can be obtained with (19) in [16]. We will refer to the free individual allocation model BGM with the BGe score

being replaced by the discrete multinomial BDe score as the Bayesian Discrete Mixture (BDM) model, and accordingly we will refer to the discrete BDe counterpart of BGM_D as the BDM_D model. We note that the BDM_D model is similar to the *nsBDe* model of Robinson and Hartemink [12], except that BDM_D leaves the inferred network structures invariant in time to allow for more information sharing among segments. We will employ the discrete counterparts BDM and BDM_D for an independent comparison of the suitability of the free allocation model and the changepoint process for inferring dynamic gene-regulatory processes (from discrete data).

2.4. MCMC Inference. We now describe an MCMC inference algorithm that can be used to obtain a sample $\{\mathcal{G}^i, \mathbf{V}^i, K^i\}_{i=1, \dots, I}$ from the posterior distribution $P(\mathcal{G}, \mathbf{V}, K | \mathbf{D})$. Our algorithm combines the structure MCMC algorithm for Bayesian networks [18] with the changepoint model (e.g., see [6, 17]), and draws on the fact that conditional on the allocation vector \mathbf{V} , separate BGe scores $P(\mathbf{D}^{(V,k)} | \mathcal{G})$ can be computed for the K data compartments. Note that this approach is equivalent to the idea underlying the allocation sampler [5]. The resulting algorithm is effectively an RJMCMC sampling scheme in the discrete space of network structures and latent allocation vectors, where the Jacobian in the acceptance criterion is always 1 and can be omitted. With probability $P = .5$, we perform a traditional structure MCMC move on the current graph \mathcal{G}^i and leave the latent vector \mathbf{V} and the number of mixture components K unchanged, symbolically: $\mathbf{V}^{i+1} = \mathbf{V}^i$ and $K^{i+1} = K^i$. A new candidate graph \mathcal{G}^{i+1} is randomly drawn out of the set of graphs $\mathcal{N}(\mathcal{G}^i)$ that can be reached from the current graph \mathcal{G}^i by deletion or addition of one single edge. The proposed graph \mathcal{G}^{i+1} is accepted with probability

$$A(\mathcal{G}^{i+1} | \mathcal{G}^i) = \min \left\{ 1, \frac{P(\mathbf{D} | \mathcal{G}^{i+1}, \mathbf{V}^i, K^i)}{P(\mathbf{D} | \mathcal{G}^i, \mathbf{V}^i, K^i)} \cdot \frac{P(\mathcal{G}^{i+1})}{P(\mathcal{G}^i)} \cdot \frac{|\mathcal{N}(\mathcal{G}^i)|}{|\mathcal{N}(\mathcal{G}^{i+1})|} \right\}, \quad (8)$$

where $|\cdot|$ is the cardinality, and the marginal likelihood terms have been specified in (6). The graph is left unchanged, symbolically $\mathcal{G}^{i+1} := \mathcal{G}^i$, if the move is not accepted. We note that the network reconstruction will be based on the marginal posterior probabilities of the individual edges, which can be estimated for each edge from the MCMC sample $\mathcal{G}^1, \dots, \mathcal{G}^I$ by the fraction of graphs in the sample that contain the edge of interest

$$\widehat{E}_{i,j} = \frac{1}{I} \sum_{k=1}^I I_{i,j}(\mathcal{G}^k), \quad (9)$$

where $I_{i,j}(\cdot)$ is the indicator function with $I_{i,j}(\mathcal{G}^k) = 1$ if there is an edge from X_i to X_j in \mathcal{G}^k .

With the complementary probability $1 - P$ we leave the graph unchanged: $\mathcal{G}^{i+1} = \mathcal{G}^i$, and perform a move on

(\mathbf{V}^i, K^i) . We change the current number of components K^i via a changepoint birth or death move, or the allocation vector \mathbf{V}^i by a changepoint reallocation move along the lines of the Reversible Jump Markov chain Monte Carlo algorithm (RJMCMC) algorithm [17].

The changepoint birth (death) move increases (decreases) K^i by 1 and may also have an effect on \mathbf{V}^i . The changepoint reallocation move leaves K^i unchanged and may have an effect on \mathbf{V}^i . If with probability $(1 - P)$ a changepoint move on (K^i, \mathbf{V}^i) is performed, we randomly draw the move type. Under fairly mild regularity conditions (ergodicity), the MCMC sampling scheme converges to the desired posterior distribution if the equation of detailed balance is fulfilled [17]. The condition of detailed balance implies that for each move a complementary move is defined, and that the acceptance probability depends on the proposal probability of the complementary move. The moves presented below are designed such that there is a unique complementary death move for each birth move and vice versa. Moreover, each reallocation move can be reversed by a single (complementary) reallocation move. The acceptance probabilities for these three changepoint moves $(K^i, \mathbf{V}^i) \rightarrow (K^{i+1}, \mathbf{V}^{i+1})$ are of the following form [17]:

$$A = \min \left\{ 1, \frac{P(\mathbf{D} | \mathcal{G}^i, \mathbf{V}^{i+1}, K^{i+1})}{P(\mathbf{D} | \mathcal{G}^i, \mathbf{V}^i, K^i)} \times R \times B \right\}, \quad (10)$$

where $R = P(\mathbf{V}^{i+1} | K^{i+1})P(K^{i+1})/P(\mathbf{V}^i | K^i)P(K^i)$ is the prior probability ratio, and B is the inverse proposal probability ratio. The exact form of these factors depends on the move type.

(i) For a changepoint reallocation (r), we randomly select one of the existing changepoints $b_j \in \{b_1, \dots, b_{K-1}\}$, and the replacement value b_j^\dagger is drawn from a uniform distribution on $[b_{j-1}, b_{j+1}]$ where $b_0 = 2$ and $b_K = m$. Hence, the proposal probability ratio is one, the prior probabilities $P(K^{i+1}) = P(K^i)$ cancel out, and the remaining prior probability ratio $P(\mathbf{V}^{i+1} | K^{i+1})/P(\mathbf{V}^i | K^i)$ can be obtained from page 720 in Green's RJMCMC paper [17]

$$R_r = \frac{(b_{j+1} - b_j^\dagger)(b_j^\dagger - b_{j-1})}{(b_{j+1} - b_j)(b_j - b_{j-1})}, \quad B_r = 1. \quad (11)$$

If there is no changepoint ($K^i = 1$) the move is rejected and the Markov chain is left unchanged.

(ii) If a changepoint birth move (b) on K^i is proposed, the location of the new changepoint b^\dagger is randomly drawn from a uniform distribution on the interval $[2, m]$; the proposal probability for this move is $b_{K^i}/(m - 2)$, where b_{K^i} is the (K^i -dependent) probability of selecting a birth move. The reverse death move, which is selected with probability $d_{(K^i+1)}$, consists in discarding randomly one of the $(K^i - 1) + 1 = K^i$ changepoints. The inverse proposal probability ratio is thus given by $B = d_{(K^i+1)}(m - 2)/(b_{K^i}K^i)$. The prior probability ratio is given by the expression at the bottom of page 720 in Green's RJMCMC paper [17] slightly modified to

allow for the fact that K components correspond to $K - 1$ changepoints, and we obtain

$$R_b = \frac{P(K^i + 1)}{P(K^i)} \frac{2K^i(2K^i + 1)}{(m - 2)^2} \frac{(b_{j+1} - b^\dagger)(b^\dagger - b_j)}{(b_{j+1} - b_j)}, \quad (12)$$

$$B_b = \frac{d_{(K^i+1)}(m - 2)}{b_{K^i}K^i}.$$

For $K^i = K^*$ the birth of a new changepoint is invalid and the Markov chain is left unchanged.

We note that the proposal probabilities b_K and $d_{(K+1)}$ for birth and death moves can be chosen as follows:

$$b_K = c \cdot \min \left\{ 1, \frac{P(K + 1)}{P(K)} \right\}, \quad (13)$$

$$d_{(K+1)} = c \cdot \min \left\{ 1, \frac{P(K)}{P(K + 1)} \right\},$$

where c is a constant that can be chosen as large as possible subject to the constraint $b_K + d_K \leq 0.9$ for $K = 1, \dots, K^*$. This choice yields both a certain acceptance rate of the MCMC sampling scheme [17] and a simple prior probability (Hastings) ratio. From $b_{K^i} \cdot P(K^i) = d_{(K^i+1)} \cdot P(K^i + 1)$ it follows that the ratio $d_{(K^i+1)}/b_{K^i}$ in the expression R_b cancels out against the prior ratio $P(K^i + 1)/P(K^i)$ in the expression B_b , and the prior probability ratio simplifies to

$$R_b B_b = \frac{2(2K^i + 1)}{(m - 2)} \frac{(b_{j+1} - b^\dagger)(b^\dagger - b_j)}{(b_{j+1} - b_j)}. \quad (14)$$

(iii) A changepoint death move (d) is the reverse of the birth move, and we obtain

$$R_d B_d = \frac{(m - 2)}{2(2K^i - 1)} \frac{(b_{j+1} - b_j)}{(b_{j+1} - b^\dagger)(b^\dagger - b_j)}. \quad (15)$$

3. Data

We have evaluated the proposed BGM_D model on various synthetic data sets. For illustration purposes, we present results obtained for two studies with small networks where the nonlinearity was implemented by a sinusoidal transformation. In a second study, we focus on the two-gene expression time series from macrophages and *Arabidopsis thaliana*, which have been used for evaluating the BGM model. More details on the experimental settings can be found in the paper on BGM [4]. In a third study, we switch from the Gaussian BGe score to the discrete multinomial BDe score to infer discrete nonstationary time series from *Drosophila melanogaster*.

3.1. Small Synthetic Networks. The first synthetic network consists of two nodes X and Y and possesses two edges. Node X has a recurrent feedback loop, symbolically $X(t) \rightarrow X(t + 1)$, and X acts as a regulator of node Y , symbolically $X(t) \rightarrow Y(t + 1)$. We consider the scenario of

a nonlinear regulatory influence that X exerts on Y , whereby we implement the nonlinearity by a sinusoid transformation of $X(t)$'s signal on $Y(t + 1)$. The state-space equations are given by

$$X(t) = X(t - 1) + c + c_X \cdot \varepsilon_X(t), \quad (16)$$

$$Y(t) = \sin(X(t - 1)) + c_Y \cdot \varepsilon_Y(t),$$

where c , c_X , and c_Y are constants, $\varepsilon_X(\cdot)$, $\varepsilon_Y(\cdot)$ are iid normally distributed random variables.

The second synthetic network is a generalisation of the two node domain where three nodes Y_1 , Y_2 , and Y_3 are regulated by X . The relationships are again realised by sinusoids, whereby we shift the periods: $Y_i(t) = \sin(X(t - 1) + \tau_i \cdot \pi) + c_Y \cdot \varepsilon_{Y_i}(t)$ with $\tau_1 = 0$, $\tau_2 = 2/3$, and $\tau_3 = 4/3$. For both networks we set the drift term $c = 2\pi/40$ to ensure that (on average) the complete period $[0, 2\pi]$ of the sinusoid is involved, and we generate $m = 41$ observations for four different combinations of the coefficients c_X and c_Y .

3.2. Bone Marrow Derived Macrophages. Interferons (IFNs) play a pivotal role in the innate and adaptive mammalian immune response against infection, and central research efforts therefore aim to elucidate their regulatory interactions [19]. For the present study, we have analysed gene expression time series from bone marrow-derived macrophages, which were sampled at $m = 25 \times 30$ minute time intervals. The macrophages were subjected to three external conditions: (1) infection with Cytomegalovirus (CMV), (2) treatment with Interferon Gamma (IFN γ), and (3) infection with Cytomegalovirus after pretreatment with IFN γ (CMV+IFN γ). Samples derived from the macrophages were hybridised to Agilent mouse genome arrays. We focus on $N = 3$ Interferon-regulatory factors Irf1, Irf2, and Irf3. These factors are the key regulators in the response of the macrophage cell to pathogens. The macrophages data sets used in the study are publicly available from <http://www.bioss.ac.uk/associates/marco/supplement/>.

3.3. Circadian Regulation in *Arabidopsis thaliana*. We have also analysed two-gene expression time series from *Arabidopsis thaliana* cells, which were sampled at $m = 13 \times 2$ hour time intervals with Affymetrix microarray chips. The expressions were measured twice independently under experimentally generated constant light condition, but differed with respect to the prehistories. In the first experimental scenario, T_{20} , the plants were entrained in a 10h:10h light/dark-cycle, while the plants in the second experimental setting, T_{28} , were entrained in 14h:14h light/dark-cycle. The analysis focuses on $N = 9$ genes, namely LHY, CCA1, TOC1, ELF4, ELF3, GI, PRR9, PRR5, and PRR3, which are known to be involved in circadian regulation [20, 21]. The *Arabidopsis* data sets used in the study are publicly available from <http://www.bioss.ac.uk/associates/marco/supplement/>.

3.4. Muscle Development in *Drosophila melanogaster*. The gene expressions in *Drosophila melanogaster* cells were sampled at $m = 67$ time-steps during four different morphological stages of life: embryonic, larval, pupal, and adult stages. Since these phases cover time periods of different lengths, gene expression profiles were collected at nonequidistant time-points. The true morphological transitions occur at time points t_{32} (embryonic to larval), t_{42} (larval to pupal), and t_{60} (pupal to adult) [22]. Like other researchers [12], we focus our analysis on $N = 13$ genes involved in muscle growth and muscle development: EVE, GFL, TWI, MLC1, SLS, MHC, PRM, ACTN, UP, MYO61F, and MSP300. The quantile-discretised binary data set of these 13 *Drosophila* genes is available from Robinson and Hartemink [12].

4. Simulations and Evaluation

We impose a fan-in restriction of size $\mathcal{F} = 3$ on the maximal number of parent nodes per node as done in related articles [2, 4]. For the Gaussian BGe model the hyperparameters of the normal-Wishart prior were chosen maximally uninformative subject to certain regularity constraints [14]. In the second study, the hyperparameters of the discrete multinomial BDe model were chosen as explained in Heckerman and Geiger [16], whereby the total prior precision parameter was set to 1.

We set $K^* = 10$, and the burn-in and the sampling-phase of MCMC runs were set to 500,000 iterations each, and we sampled every 1,000 iterations during the sampling-phase. For each data set, we started 5 independent MCMC simulations from different initialisations, and we computed the potential scale reduction factor PSRF [23] based on the marginal edge posterior probabilities. As we observed, a sufficient degree of convergence for all data sets (PSRF < 1.2), we report only the results of the empty-seeded (graphs without any edges) MCMC runs. We note that each single MCMC simulation (even for the bigger domains *Arabidopsis* with $N = 9$ genes and *Drosophila* with $N = 13$ genes) was accomplished within a few hours using Matlab[©] code on a SunFire X4100M2 machine with MAD Opteron 2224 SE dual-core processor.

More generally, we note that the computational complexity in network structure space is $N \sum_{k=0}^{\mathcal{F}} \binom{N}{k} \sim O(N^{\mathcal{F}+1})$, where N is the number of nodes and \mathcal{F} is the maximum number of parent nodes. The complexity in changepoint configuration space is of order m^K , where K is the number of changepoints, and m is the number of time points. Hence, the problem is of polynomial complexity in N and m , and *not* exponential complexity (i.e., it is not NP-hard). Polynomial complexity—as opposed to exponential complexity—does not impose any principled restrictions on the network size. However, the practical feasibility will depend on varying factors, like the efficiency of the software implementation and the capacity of parallel clusters. It also depends on the information content of the data, as increasing the number of nodes with limited number of experimental replications will increase the intrinsic uncertainty of inference. The practical decision on how many nodes and how large a network to

consider will therefore usually be based on some preliminary data analysis.

We note that we allowed for self-feedback loops for the synthetic data only. For the three real applications to biological systems, we ruled out self-feedback loops altogether to enable direct comparability with the results reported for the BGM model [4].

If two independent data sets \mathbf{D} and \mathbf{D}_{\ddagger} are available for a network domain, predictive probabilities of the models BGM and BGM_D can be estimated straightforwardly [4]. For example, having sampled $\{\mathcal{G}^i, \mathbf{V}^i, K^i\}_{i=1, \dots, I}$ from the posterior distribution $P(\mathcal{G}, \mathbf{V}, K | \mathbf{D})$ via the MCMC inference scheme described above, the predictive probability of BGM_D can be estimated by

$$P(\mathbf{D}_{\ddagger} | \mathbf{D}, \text{BGM}_D) = \frac{1}{I} \sum_{i=1}^I \prod_{k=1}^{K^i} \Psi(\mathcal{G}^i, \mathbf{D}_{\ddagger}^{(\mathbf{V}^i, k)} | \mathbf{D}^{(\mathbf{V}^i, k)}), \quad (17)$$

where

$$\begin{aligned} & \Psi(\mathcal{G}^i, \mathbf{D}_{\ddagger}^{(\mathbf{V}^i, k)} | \mathbf{D}^{(\mathbf{V}^i, k)}) \\ &= \prod_{n=1}^N \int \prod_{t: \mathbf{V}^i(t)=k} P(X_n(t) = \mathbf{D}_{\ddagger n, t} | \pi_n(t-1) = \mathbf{D}_{\ddagger(\pi_n, t-1)}, \boldsymbol{\theta}_n) \\ & \quad \times P(\boldsymbol{\theta}_n | \pi_n, \mathbf{D}^{(\mathbf{V}^i, k)}) d\boldsymbol{\theta}_n. \end{aligned} \quad (18)$$

That is, when computing the BGe score for the compartment $D^{(\mathbf{V}^i, k)}$ with (6) and (7), the local prior distributions $P(\boldsymbol{\theta}_n | \pi_n)$ in (7) are replaced by posterior distributions $P(\boldsymbol{\theta}_n | \pi_n, \mathbf{D}^{(\mathbf{V}^i, k)})$. This results in a straightforward modification of the BGe score as follows. In (13) in Geiger and Heckerman [14], those training data that have been allocated to component k , symbolically $\mathbf{D}^{(\mathbf{V}^i, k)}$, are included in the conditioning part of the distribution, and the sufficient statistics are adjusted accordingly.

5. Results

5.1. Inference on Synthetic Data. For the synthetic sinusoid data, the true underlying network topologies are known so that the network reconstruction accuracy can be assessed via the area under the ROC (receiver operator characteristic) curve: AUC; this is a standard criterion that has been applied in numerous related articles (e.g., [24]). Figure 1 shows histograms of the average AUC scores obtained from the marginal edge posterior probabilities for the sinusoid networks with $N = 2$ and $N = 4$ nodes. Figure 1 is laid out as matrix, in which the rows and columns correspond to different c_X and c_Y coefficients (noise levels). We note that an increase of c_X reduces the autocorrelation of X , while increasing c_Y blurs the functional dependence of $Y(t+1)$ on $X(t)$. The autocorrelation of Y is jointly influenced by both parameters. From the histograms it can be seen that the novel BGM_D model leads to a better network reconstruction accuracy in terms of average AUC values than the standard

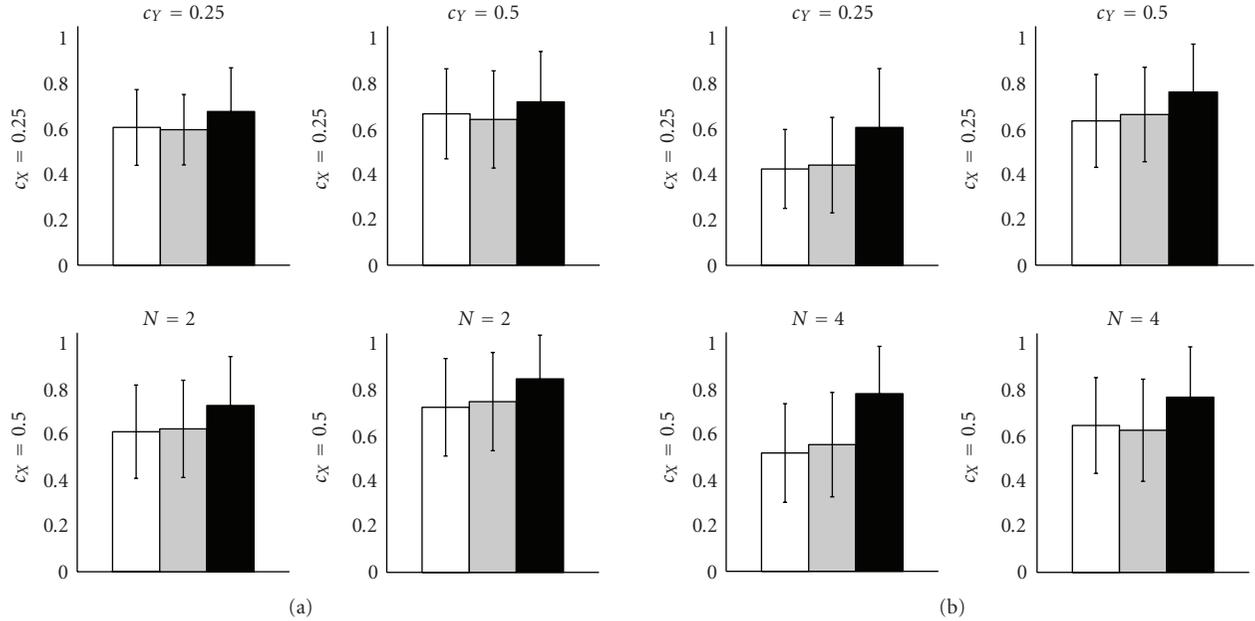


FIGURE 1: AUC histograms—Cross-method comparison on synthetic sine data. AUC scores for the synthetic sine network data with $N = 2$ nodes (a) and $N = 4$ nodes (b). The figure is laid out as a matrix, where rows and columns correspond to different noise levels c_X (rows) and c_Y (columns). In each histogram, the white bar shows the average AUC score for the BGe model. The grey bar shows the average AUC score of the BGM model, and the black bar shows the AUC score for the proposed BGM_D model. Each histogram shows averages and standard deviations obtained from 50 independent data instantiations.

BGe model and the BGM model in the majority of scenarios. Further investigations showed that BGe and BGM yield lower AUC scores, since they tend to infer spurious self-feedback loops on node Y ($N = 2$) or on the nodes Y_i ($N = 4$), respectively.

This trend can be visualised by histograms of the average edge posterior probabilities. As an example, Figure 2 shows the average marginal edge posterior probabilities of the four possible edges for the $N = 2$ nodes sinusoid network with $c_X = 0.5$ and $c_Y = 0.5$. Consistently, all three models under comparison assign the highest posterior probability to the true self loop $X \rightarrow X$ and the lowest posterior probability to the false edge $Y \rightarrow X$. But BGe and BGM favour the spurious feed-back loop $Y \rightarrow Y$ over the true edge $X \rightarrow Y$ while the proposed BGM_D suppresses the false self-feedback loop and assigns a higher edge posterior probability to the true edge $X \rightarrow Y$. This shows that BGM_D yields a higher network reconstruction accuracy (see Figure 1), as it is less susceptible to inferring spurious self-feedback loops (see Figure 2).

5.2. Inference on Macrophages Data. For the macrophages data the BGM model inferred a biologically plausible state change in the host macrophage brought about by infection (CMV) or immune activation (IFN γ), and a less pronounced state change in the combined condition CMV+IFN γ [4]. We compare these findings with results obtained with the novel BGM_D model. The fraction of sampled states for which two time points t_i and t_j are allocated to the same component k ($1 \leq k \leq K$) can be used as a connectivity

measure $C(t_i, t_j)$, and the resulting temporal connectivity structures are displayed graphically as heat maps in Figure 3. All six heatmaps in Figure 3 reflect the two-stage nature of the gene-regulatory processes in the host macrophages: the first part (time points (t_2, \dots, t_6)) and the last part of the three time series (time points (t_7, \dots, t_{25})) are allocated to different components. For all three conditions, a stronger separation between the two regulatory states is inferred by the BGM_D model (see Figures 3(d), 3(e), and 3(f)). It appears that the BGM_D inference results are more consistent, as even for the combined condition (CMV+IFN γ) a clear trend towards a dichotomous regulatory process can be found (see Figure 3(d)). This finding (stronger separation) is consistent with our conjecture that the novel BGM_D assigns neighbouring time-points to the same compartment more likely a priori. Interestingly, the BGM inference outlier at time point t_9 in Figure 3(b) yields a certain trend for a subdivision of the second compartment (t_7, \dots, t_{25}) by the BGM_D model. Instead of one outlying time point two substages (t_7, \dots, t_{10}) and (t_{11}, \dots, t_{25}) are inferred (see Figure 3(e)). To provide statistical evidence that the new BGM_D model does not overfit the data, we compute predictive probabilities for the BGM_D model and compare them with those reported for the BGM model [4].

To this end, we treat the three experiments as independent replications, and check via predictive probabilities whether the superiority of the proposed BGM_D model can be confirmed statistically. Table 1 gives the predictive probabilities reported in the BGM paper [4] and those obtained with the proposed BGM_D model. As the predictive probabilities

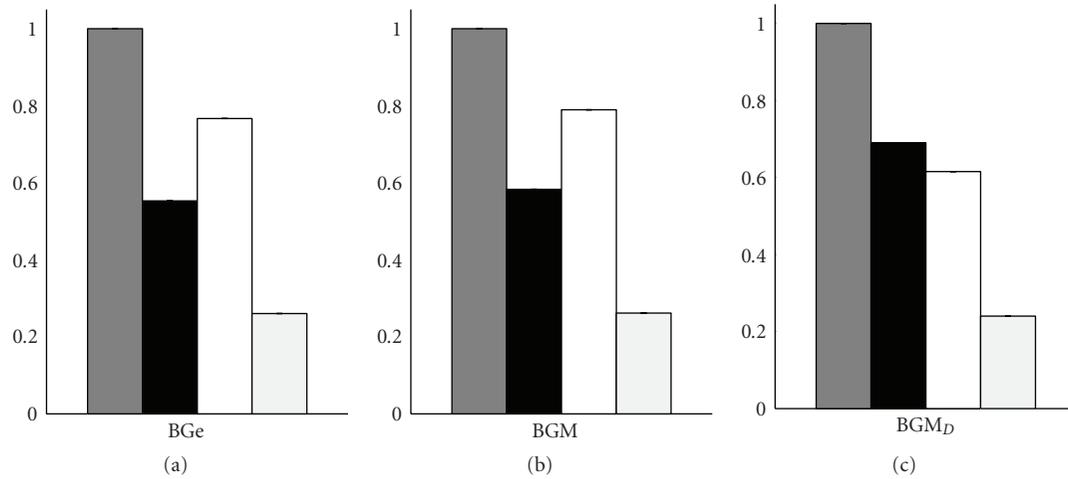


FIGURE 2: Edge Posterior Probabilities—Cross-method comparison on synthetic sine data. The figure shows three histograms of the inferred marginal edge posterior probabilities in the sinusoid network with $N = 2$ nodes and $c_X = 0.5$ and $c_Y = 0.5$ as obtained with BGe (a), BGM (b), and BGM_D (c). In each histogram, the four bars represent the four possible edges: Left: self-loop $X \rightarrow X$ (true); centre left: $X \rightarrow Y$ (true); centre right: self-loop $Y \rightarrow Y$ (false); right: $Y \rightarrow X$ (false). Each bar shows the average marginal posterior probability, averaged over 50 independent data instantiations. It is seen that BGe and BGM have a high propensity for learning the spurious feedback loop $Y \rightarrow Y$ (centre right white bars). BGM_D (right histogram) assigns a higher posterior probability to the correct edge $X \rightarrow Y$ (centre left black bar) and suppresses the spurious feedback loop $Y \rightarrow Y$ (centre right white bar)

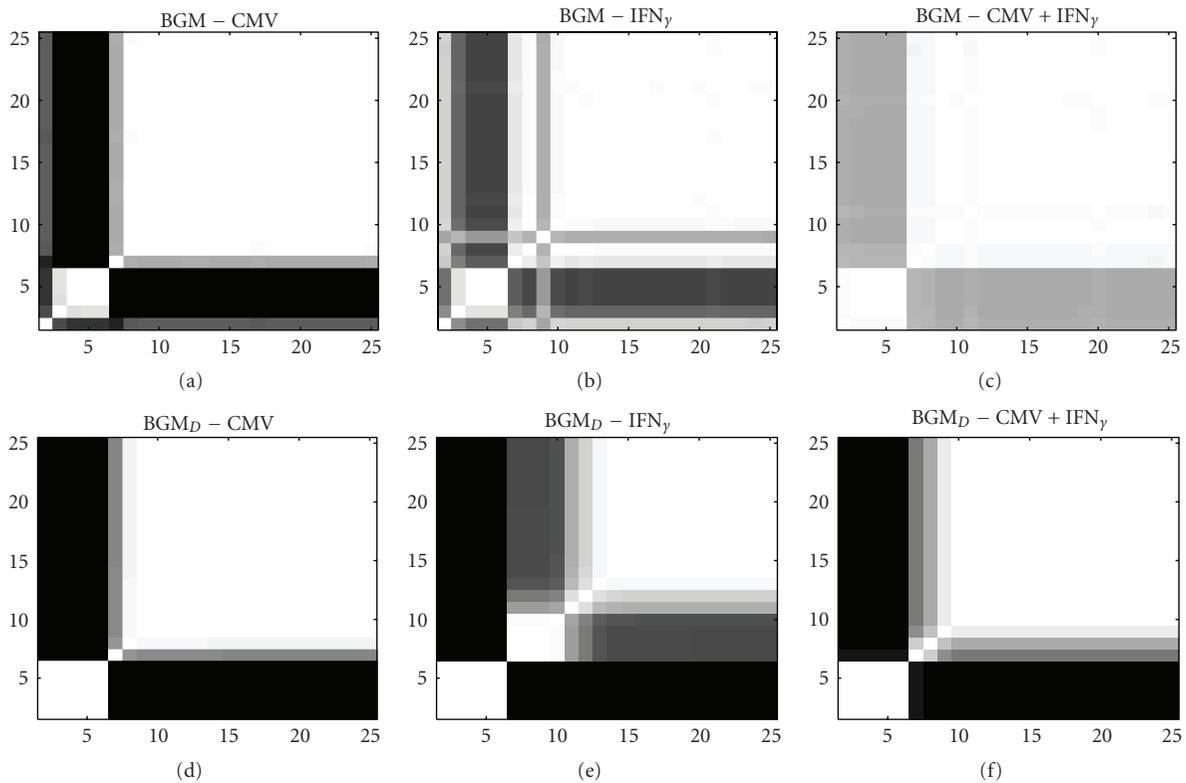


FIGURE 3: Heat maps—macrophages data. Graphical heat map presentation of the temporal connectivity structure for the macrophage gene expression time series. (a), (b), and (c): Heat matrices for experiments CMV (a), IFN_γ (b), and CMV+IFN_γ (c) inferred with the BGM model. (d), (e), and (f): Heat matrices for experiments CMV (d), IFN_γ (e), and CMV+IFN_γ (f) inferred with the novel BGM_D model. Each heat map indicates the estimated posterior probability of two time points being assigned to the same compartment (mixture component). The probabilities are represented by a grey shading, where white corresponds to a probability of 1, and black corresponds to a probability of 0. The numbers on the axes represent the time points of the time course experiment.

TABLE 1: Predictive Probabilities: Macrophages data. Logarithmic predictive probabilities for the macrophage data: $\log_e(P(\mathbf{D}_\ddagger|\mathbf{D}))$ for BGe, BGM (as reported earlier [4]) and the new BGM_D model. The standard deviations of the logarithmic probabilities are given in brackets. We note that the BGM values could be confirmed by our reanalysis of the data: the deviations were smaller than one standard deviation.

$\mathbf{D} = \mathbf{D}_{\text{TRAIN}}$	Model	CMV	$\mathbf{D}_\ddagger = \mathbf{D}_{\text{TEST}}$	
			IFN γ	CMV and IFN γ
CMV	BGe	—	-76.01 (± 0.07)	-45.26 (± 0.03)
	BGM	—	-63.63 (± 0.02)	-33.80 (± 0.38)
	BGM_D	—	-59.13 (± 0.02)	-31.76 (± 0.25)
IFN γ	BGe	-56.78 (± 0.05)	—	-57.30 (± 0.05)
	BGM	-39.62 (± 0.02)	—	-42.69 (± 0.11)
	BGM_D	-34.08 (± 0.14)	—	-39.11 (± 0.10)
CMV+IFN γ	BGe	-37.76 (± 0.08)	-69.19 (± 0.06)	—
	BGM	-21.67 (± 0.33)	-53.26 (± 0.51)	—
	BGM_D	-18.58 (± 0.09)	-51.11 (± 0.21)	—

for BGM_D are systematically better than those of BGM, we conclude that the BGM_D model yields more stable inference results, that is, a better generalisation performance.

5.3. Inference on *Arabidopsis thaliana* Data. For the *Arabidopsis thaliana* data, the BGM model also inferred a biologically plausible two-stage process [4]. In this application, the two stages are likely to be related to the diurnal nature of the dark-light cycle influencing the circadian genes. The plants were subjected to different prehistories, related to different lengths of the artificial, experimentally controlled light-dark cycle. The plants in experimental scenario T_{28} were entrained in an increased day length of 14 hours light followed by 14 hours darkness and in experiment T_{20} the plants were entrained in a decreased day length of 10 hours light followed by 10 hours darkness. As an effect of these two entrainments, a phase shift in the gene-regulatory processes between these two experiments was expected [4]. The BGM model inferred a certain trend for a phase shift of the changepoint (subjective day to subjective night) of about 4–6 hours as a consequence of the increased day length. The heat maps in Figures 4(a) and 4(b) show that the connected blocks (compartments) of the time series are shifted along the diagonal by 2-3 time-points (4–6 hours). The BGM_D model infers the same trend but with a stronger separation score between these compartments (see Figures 4(c) and 4(d)). We note that the BGM_D model is based on changepoints so that compartments once left cannot be revisited. That is, while the BGM model tends to allocate the first time points (t_2, t_3) and the last time points (t_9, \dots, t_{13}) in experiment T_{28} to one single component (light grey shading in the top right and bottom left area of the heat map in the top centre panel of Figure 4), the BGM_D model has to allocate the last time points (t_9, \dots, t_{13}) to an additional third component, as the first compartment (t_2, t_3) cannot be reused after the transition to the second compartment (t_4, \dots, t_8).

As for the macrophages data, predictive probabilities can be computed by treating the two experiments T_{20} and T_{28} as independent replications. Table 2 gives the predictive probabilities reported for the BGM model [4] and those obtained with the new BGM_D model. In consistency with the

results for the macrophages data, the resulting two predictive probabilities for BGM_D are better than those of the BGM model. A scatter plot $\widehat{E}_{i,j}(T_{20})$ versus $\widehat{E}_{i,j}(T_{28})$ of the inferred (marginal) posterior probabilities of the individual edges for the BGM_D model inference is given in Figure 5. As the individual edge posterior probabilities $E_{i,j}$ do not differ drastically, we extract a network structure from the averaged probabilities: $\widehat{E}_{i,j} = (\widehat{E}_{i,j}(T_{20}) + \widehat{E}_{i,j}(T_{28}))/2$.

Figure 6 shows the gene interaction network that is predicted when keeping all edges with marginal posterior probability above 0.5. There are two groups of genes. Empty circles in the figure represent morning genes (i.e., genes whose expression peaks in the morning), shaded circles represent evening genes (i.e., genes whose expression peaks in the evening). There are several directed edges pointing from the group of morning genes to the evening genes such that each evening gene is regulated by at least one morning gene. Moreover, the two genes LHY and CCA1 seem to play a central role. This result is consistent with biological findings, where the morning genes were found to activate the evening genes, with LHY and CCA1 being central regulators [25]. Our reconstructed network also contains edges pointing into the opposite direction, from the evening genes back to the morning genes. This finding is consistent with biological observations [25], where the evening genes were found to inhibit the morning genes via negative feedback loops. In the reconstructed network, there are 9 edges (drawn in black) originating from the four morning genes while only 7 edges (drawn in grey) originate from the group of five evening genes. Biologically, this means that the activity of the morning genes is stronger than the activity of the group of evening genes and that the regulatory mechanisms are dominated by the morning genes in the network topology. This finding is consistent with the fact that following the light-dark cycle entrainment, the experiments were carried out in constant-light condition, resulting in a higher activity of the morning genes overall. Within the group of evening genes, the reconstructed network contains an edge between GI and TOC1. This interaction has been confirmed independently [26]. Hence, while a proper evaluation of the reconstruction accuracy is currently unfeasible—like

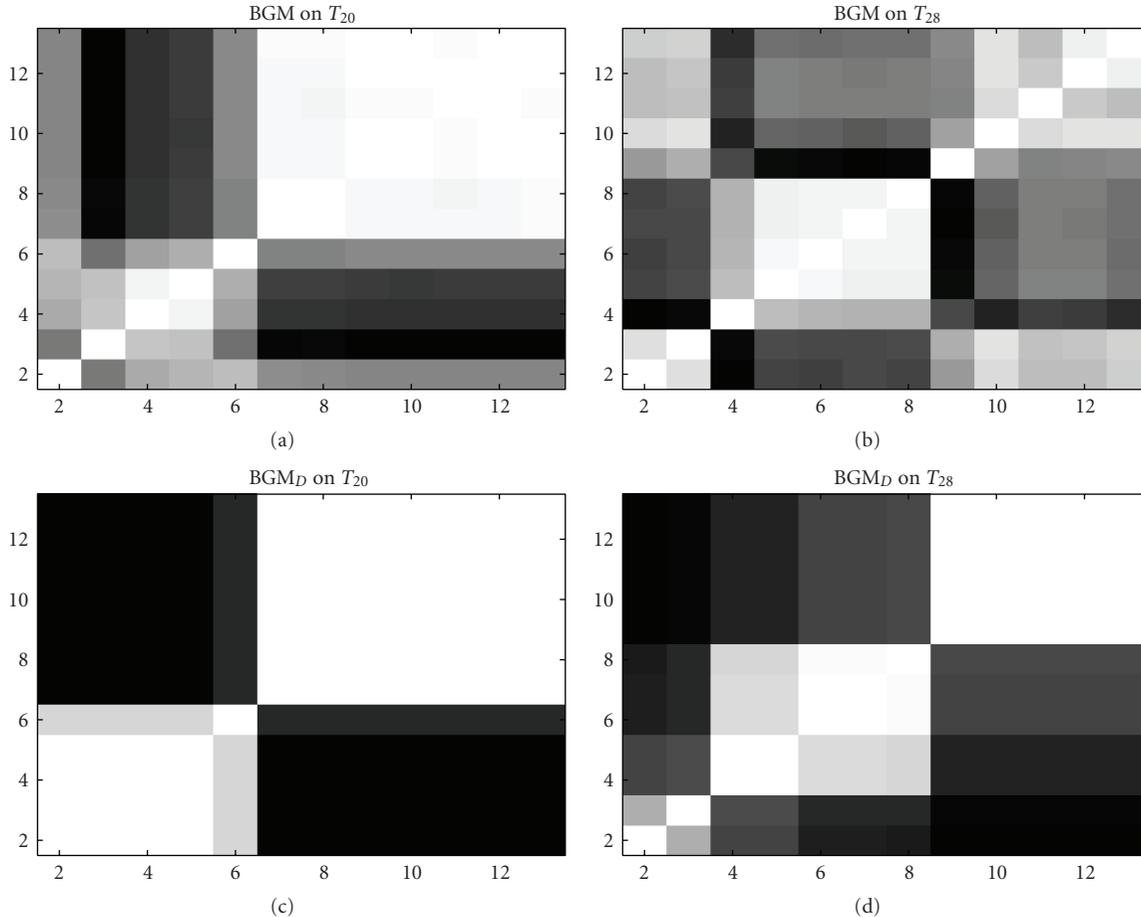


FIGURE 4: Heat maps. *Arabidopsis* data. Graphical heat map representations of the temporal connectivity structures for the *Arabidopsis thaliana* data. (a) and (b): heat matrices for experiments T_{20} (a) and T_{28} (b) inferred with the BGM model. (c) and (d): heat matrices for experiments T_{20} (c) and T_{28} (d) inferred with the novel BGM_D model. Each heat map indicates the posterior probability of two time points being assigned to the same compartment (mixture component). The probabilities are represented by a grey shading, where white corresponds to a probability of 1, and black corresponds to a probability of 0. The numbers on the axes represent the time points of the time course experiment.

many related studies, we lack a gold-standard owing to the unknown nature of the true interaction network—our study suggests that the essential features of the reconstructed network are biologically plausible and consistent with the literature.

5.4. Inference on *Drosophila melanogaster* Data. For an independent comparison of the free allocation model and the changepoint model, we carried out an analysis similar to Robinson and Hartemink [12] on the binary *Drosophila* muscle development gene expression time series. This time series can be analysed with the discrete counterparts BDM and BDM_D of our Gaussian Mixture models; see Section 2.3 for details. The graphical heat map representations in Figure 7 show that the BDM model does not infer the morphological stages of *Drosophila melanogaster*. Almost all time-points are strongly connected (white shading) and no separated blocks of connected time points have been inferred. Only a few time-points are allocated (as outliers)

TABLE 2: Predictive probabilities: *Arabidopsis thaliana* data. Logarithmic predictive probabilities for the *Arabidopsis thaliana* data: $\log_e(P(\mathbf{D}_\dagger|\mathbf{D}))$ for BGe, BGM (as earlier reported [4]) and the new BGM_D model. The standard deviations of the logarithmic probabilities are given in brackets. We note that the BGM values could be confirmed by our reanalysis of the data: the deviations were smaller than one standard deviation.

\mathbf{D}	H_i	$\mathbf{D}_\dagger = T_{20}$	$\mathbf{D}_\dagger = T_{28}$
T_{20}	BGe	—	$-64.29 (\pm 0.29)$
	BGM	—	$-53.69 (\pm 0.42)$
	BGM_D	—	$-52.39 (\pm 0.78)$
T_{28}	BGe	$-63.93 (\pm 0.22)$	—
	BGM	$-54.78 (\pm 0.63)$	—
	BGM_D	$-48.69 (\pm 0.32)$	—

to other mixture components (black vertical and horizontal lines).

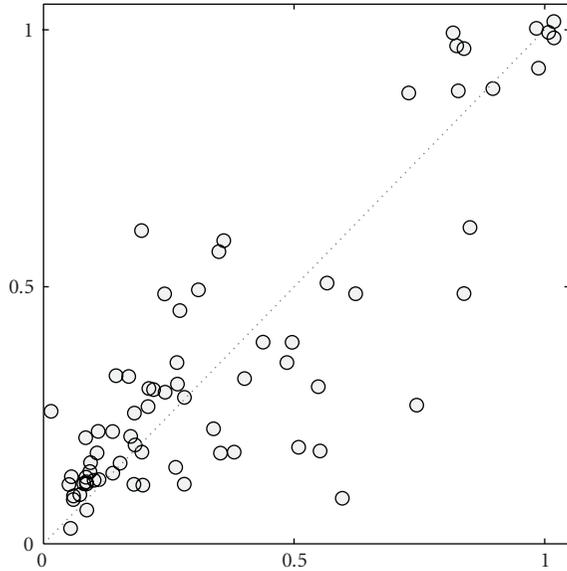


FIGURE 5: BGM_D scatter plot. Arabidopsis data. Scatter plot of the marginal edge posterior probabilities inferred with the proposed BGM_D model. In the plot the marginal edge posterior probabilities for time series T_{20} : $\widehat{E}_{i,j}(T_{20})$ (horizontal axis) are plotted versus the marginal edge posterior probabilities for time series T_{28} : $\widehat{E}_{i,j}(T_{28})$ (vertical axis).

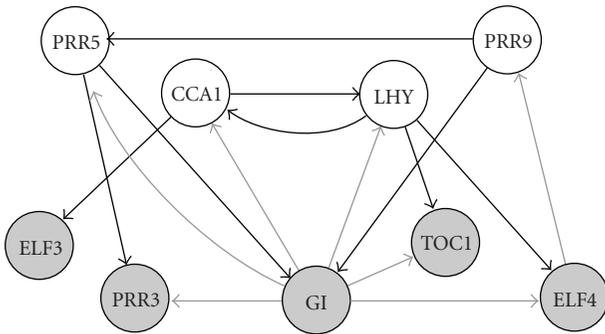


FIGURE 6: Predicted regulatory network of nine circadian genes in *Arabidopsis thaliana*. From the averaged marginal edge posterior probabilities (average of $\widehat{E}_{i,j}(T_{20})$ and $\widehat{E}_{i,j}(T_{28})$) of the proposed BGM_D model inference results a regulatory network can be extracted. Empty circles represent morning genes. Shaded circles represent evening genes. Edges indicate predicted interactions with an inferred marginal posterior probability greater than 0.5. Edges are black (grey) if they refer to a morning (evening) gene as regulator.

The novel BGM_D infers a clear block structure with different separated compartments. The connectivity structure corresponds well to the first two morphological transitions (i) embryonic to larval ($t_{31} \rightarrow t_{32}$) and (ii) larval to pupal ($t_{41} \rightarrow t_{42}$), whereby the separation between the embryonic and the larval stage is less pronounced (grey shading) than the separation between the larval and the pupal stage (black shading). The exact third morphological transition pupal to adult ($t_{59} \rightarrow t_{60}$) has not been inferred

but it can be seen that two changepoints occur during the pupal stage before the third transition to the adult stage. Figure 8 shows a graphical presentation of the changepoint location posterior probabilities inferred with BGM_D and the same trends become obvious: the first two stage transitions have been inferred correctly while two further changepoints occur before the morphological stage transition from pupal to adult. We note that Robinson and Hartemink did find the same trends with their *nsBDe* model and they conclude that the third premature transition in the gene regulatory process can be explained biologically, since the gene expression program governing the transition from pupal to adult morphology should be active well before the time of the real morphological transition [12].

6. Discussion

Our empirical results have shown that the proposed BGM_D model performs consistently better than the BGM model. A possible explanation could be related to the latent space complexity of the models. For two components, we have $m - 2$ changepoint locations, but 2^{m-1} free allocations. This difference in latent space complexity gets aggravated for more components. The difference in performance between BGM_D and BGM could therefore be a consequence of the different degrees of convergence of the MCMC simulations. However, our convergence diagnostics based on potential scale reduction factors [23] did not indicate any significant difference in the convergence. It therefore appears that the higher latent space complexity of the BGM model does not adversely affect the MCMC convergence for gene expression time series of the length investigated in our study. This suggests that another explanation for the better performance of BGM_D over BGM has to be found.

We will discuss that the performance difference is most likely a consequence of the different prior probabilities intrinsic to the models, which determine the factor R in (10). Since both models BGM and BGM_D employ the same Poisson distribution with parameter $\lambda = 1$ truncated to the interval $[1, K^*]$ as prior for the number of mixture components, the difference between the two models BGM and BGM_D is imposed by the prior distribution of the allocation vector $P(\mathbf{V}|K)$. While the BGM model is based on a free allocation, the BGM_D model takes the time structure into account and employs a changepoint process. In this section we describe the differences between the two models in detail by three theoretical examinations, and these examinations reveal trends that appear to be immediately related to some of our empirical findings.

In the first study, we compare the (temporal) connectivity structures that are introduced by the prior $P(\mathbf{V}|K)$. To this end, we infer the prior distribution $P(K)P(\mathbf{V}|K)$ of both models BGM and BGM_D via MCMC simulations that are purely prior-driven. That is, we employ an empty data set (without any data points) so that all acceptance probabilities depend on the prior probability ratios only, as $P(\mathbf{D} = \emptyset | (\mathcal{G}, \mathbf{V}, K)) = 1$ for all combinations of \mathcal{G} , \mathbf{V} , and K . Note that sampling network structures via edge

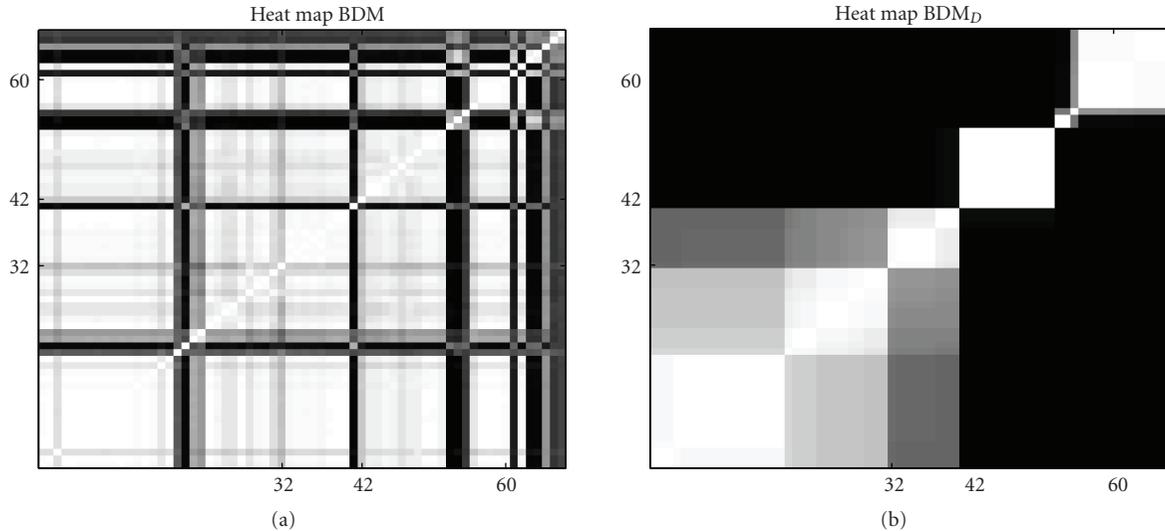


FIGURE 7: Heat maps for the discrete *Drosophila* data. Graphical heat map representations of the temporal connectivity structures for the *Drosophila melanogaster* data. (a) Heat matrix inferred with the BDM model. (b) Heat matrix inferred with the novel changepoint variant BMD_D . Each heat map indicates the posterior probability of two time points being assigned to the same compartment (mixture component). The probabilities are represented by a grey shading, where white corresponds to a probability of 1, and black corresponds to a probability of 0. The numbers on the axes represent the time points of the time course experiment. Both axes have been ticked at the three real morphological stage transitions: embryonic to larval ($t_{31} \rightarrow t_{32}$), larval to pupal ($t_{41} \rightarrow t_{42}$), and pupal to adult ($t_{59} \rightarrow t_{60}$).

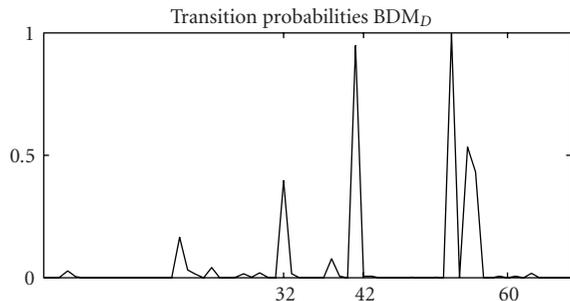


FIGURE 8: BMD_D transition posterior probabilities for *Drosophila* data. Graphical representation of the posterior probabilities of the changepoint locations inferred with the novel BMD_D . The transition posterior probabilities (vertical axis) are plotted against the time axis (horizontal axis). The time axis has been ticked at the three real morphological stage transitions: embryonic to larval ($t_{31} \rightarrow t_{32}$), larval to pupal ($t_{41} \rightarrow t_{42}$), and pupal to adult ($t_{59} \rightarrow t_{60}$). We note that the BDM model is based on a free individual allocation of time points so that a transition posterior probability plot cannot be interpreted properly.

operation moves on the graph \mathcal{G} becomes obsolete, because (i) the graphs do not have any effect on the likelihoods for an empty data set, and (ii) the graph priors cancel out in the prior probability ratio if a uniform graph prior $P(\mathcal{G}) = \text{const}$ is used. We set $m = 26$; note that in this theoretical consideration based on empty data, m only determines the length of the allocation vector.

After running 10 independent MCMC simulations with $m = 26$ to infer the prior distribution $P(K) \cdot P(\mathbf{V}|K)$, we can compute the average prior connectivity strengths from

the sampled allocation vectors. As before, the fraction of sampled allocation vectors for which two time points t_i and t_j are allocated to the same component k ($1 \leq k \leq K$) can be used as a connectivity measure $C(t_i, t_j)$. Figure 9 shows heat maps of the inferred prior connectivity structure for the free allocation BGM model and the proposed changepoint process BGM_D model. The heat maps confirm our earlier conjecture that the proposed BGM_D model, which takes the time structure of the data into account, allocates neighbouring time points to the same compartment more likely a priori than the BGM model. More precisely, it can be seen from Figure 9(a) that the prior connectivity strength $C(t_i, t_j)$ is the same for all t_i and t_j with $t_i \neq t_j$ in the BGM model. On the contrary, for the BGM_D model (Figure 9(b)), the connectivity strength $C(t_i, t_j)$ decreases with the temporal distance between t_i and t_j : for three time points t_i, t_j and t_k with $t_i < t_j < t_k$ we have $C(t_i, t_j) < C(t_i, t_k)$. This finding explains why the proposed BGM_D model yields a stronger separation between the light:darkness induced stages in *Arabidopsis thaliana* (see Figure 4): the two-stage structure of gene-regulation in *Arabidopsis* is of a temporal form that is supported by the allocation vector prior $P(\mathbf{V}|K)$ of the BGM_D model.

In the second and third theoretical study, we cross-compare the BGM model and the proposed BGM_D model in terms of prior probability ratios between the heterogeneous (nonstationary) and the homogeneous (stationary) state. That is, both models can be in the homogeneous state (i.e., the complete time series is modelled as one single compartment ($K = 1$)) or in a heterogeneous state (i.e., the time series is divided into K different compartments ($K > 1$)). For the BGM as well as the BGM_D model

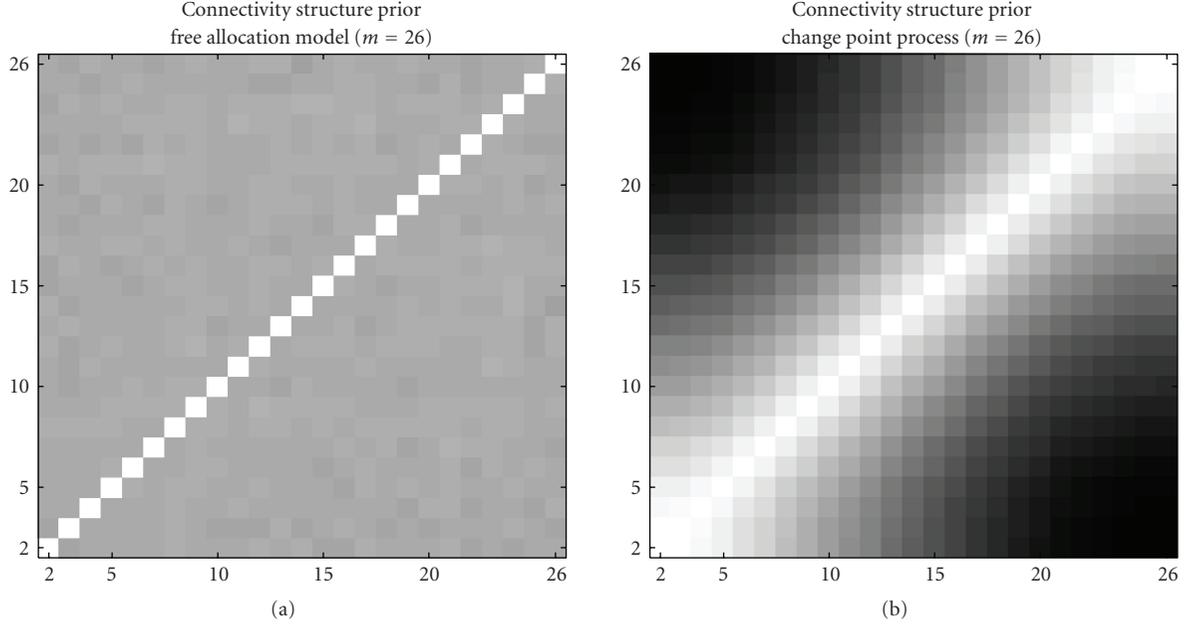


FIGURE 9: Prior connectivity structure for time series of length $m = 26$. Graphical heat map representations of the temporal connectivity structures imposed by the prior distribution $P(K) \cdot P(\mathbf{V} | K)$ for $m = 26$ time points. (a) Free allocation model (BGM). (b) Change point process (BGM_D). Each heat map indicates the prior probability of two time points being assigned to the same compartment. The probabilities are represented by a grey shading, where white corresponds to a probability of 1, and black corresponds to a probability of 0.5. The connectivity strengths were estimated from 10 independent MCMC simulations. In these simulations, an empty data set (without any data points) was used so that the inference was driven exclusively by the prior probability distribution $P(K) \cdot P(\mathbf{V} | K)$.

the prior probability ratio R between (i) the heterogeneous state consisting of two segments ($K = 2$) and (ii) the homogeneous state ($K = 1$) is given by

$$R = \frac{P(K = 2)}{P(K = 1)} \cdot \frac{P(\mathbf{V} | K = 2)}{1}, \quad (19)$$

since $P(\mathbf{V} | K = 1) = 1$ for both models. We consider the scenario where the time series is of length m and the allocation vector \mathbf{V} divides the time series into two (non-empty) connected segments t_2, \dots, t_j and t_{j+1}, \dots, t_m ($2 \leq j \leq m$). The prior probability ratio of the BGM model is then given by

$$R_{\text{BGM}} = \frac{P(K = 2)}{P(K = 1)} \cdot \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + (m - 1))} \cdot \prod_{k=1}^2 \frac{\Gamma(\alpha_k + m_k)}{\Gamma(\alpha_k)}, \quad (20)$$

where $\alpha_1 = \alpha_2 = 1$, $\alpha_0 = \alpha_1 + \alpha_2 = 2$, and m_k is the number of time points that have been allocated to the k th segment ($k = 1, 2$); that is, $m_1 = j - 1$ and $m_2 = m - j$ in our scenario. See [4] for details. The proposed BGM_D model requires a changepoint in the interval $[j, j + 1]$ and it can be derived straightforwardly from (12):

$$R_{\text{BGM}_D} = \frac{P(K = 2)}{P(K = 1)} \cdot \int_j^{j+1} \frac{6(m - b_1)(b_1 - 2)}{(m - 2)^3} db_1. \quad (21)$$

In the second theoretical study we vary the length of the time series $m = 3, 5, 7, \dots, 25$, and we consider a heterogeneous time series consisting of two equally-spaced segments

$t_2, \dots, t_{\lfloor m/2 \rfloor + 1}$ and $t_{\lfloor m/2 \rfloor + 2}, \dots, t_m$. This corresponds to $m_1 = m_2 = 0.5 \cdot (m - 1)$ in the BGM model. For the BGM_D model, we obtain with $j = \lfloor m/2 \rfloor + 1$ that the changepoint has to be located in the interval $b_1 \in [t_{\lfloor m/2 \rfloor + 1}, t_{\lfloor m/2 \rfloor + 2}]$. Figures 10(a) and 10(b) show the resulting (logarithmic) prior probability ratios in dependence on m . It can be seen that the prior ratio R for the BGM model is considerably lower than for the BGM_D model. Moreover, the logarithmic plot in Figure 10(b) shows that the prior ratio of the BGM model shows a much stronger decrease with the sample size m than the BGM_D model. This suggests that the BGM model imposes a more severe penalty for complexity (non-stationarity), which increases with increasing sample size m . This tendency may explain the finding in [4] for the macrophage gene expression time series, which we have reproduced in the present study (Figure 3(c)): the BGM model does not infer a clear two-phase nature of the time series under simultaneous immune activation (with IFN γ) and viral infection (with CMV). A possible biological explanation was offered in [4]. However, the novel BGM model does not support the hypothesis of a decreased probability for the two-phase nature (Figure 3(f)). Moreover, the previous analysis has revealed that a strong penalty against the two-phase process is inherent in the BGM model. This suggests that the results reported in [4], which we have reproduced in our study, might be an artefact of the BGM model rather than of genuine biological nature.

In the third theoretical study, we fix the length of the time series $m = 26$ and we vary the last time point $j = 2, 3, \dots, 25$

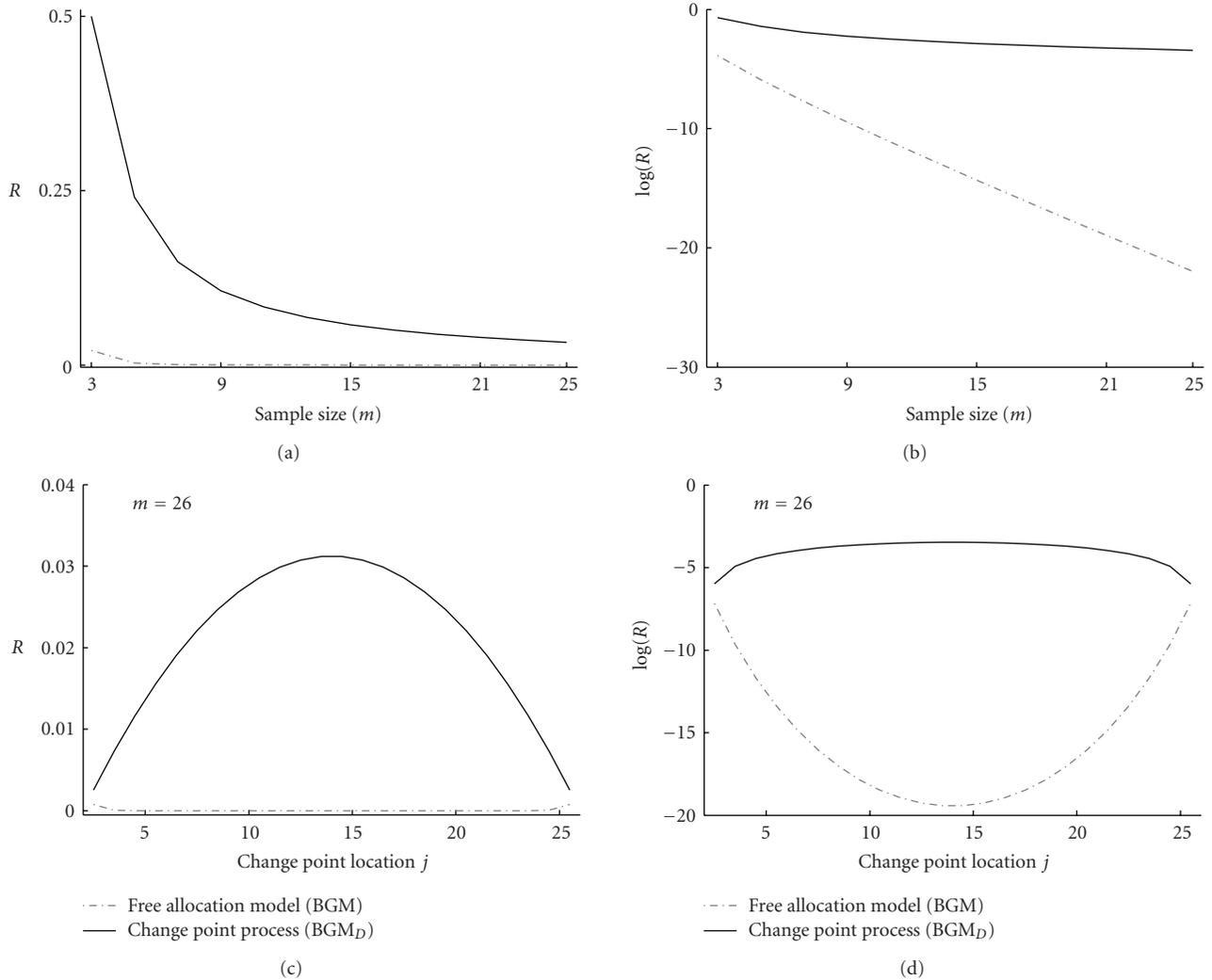


FIGURE 10: Prior probability ratios between the heterogeneous and the homogeneous state for (i) varying time series length m ((a) and (b)) and (ii) varying segment length proportions ((c) and (d)). (a) and (b): prior probability ratio R between (i) the heterogeneous state that consists of two equally-spaced segments $t_2, \dots, t_{\lfloor m/2 \rfloor + 1}$ and $t_{\lfloor m/2 \rfloor + 2}, \dots, t_m$ and (ii) the homogeneous state consisting of one single segment t_2, \dots, t_m . The prior probability ratios (vertical axis) are plotted in dependence on the time series length $m = 3, 5, 7, \dots, 25$ (horizontal axis). The prior probability ratio was defined in (19). For clarity the logarithmic prior probability ratios are plotted in (b). (c) and (d): prior probability ratio R for a time series of length $m = 26$ between (i) the heterogeneous state with two segments t_2, \dots, t_j and t_{j+1}, \dots, t_m and (ii) the homogeneous state consisting of one single segment t_2, \dots, t_m only. The prior probability ratios (vertical axis) are plotted in dependence on the changepoint location (horizontal axis). For the sake of clarity the logarithmic prior probability ratios are plotted in (b). See text for further details.

of the first segment in the heterogeneous state to illustrate the effect of unequal segment lengths $j - 1$ (t_2, \dots, t_j) and $m - j$ (t_{j+1}, \dots, t_m). For the BGM model, this scenario yields $m_1 = j - 1$ and $m_2 = m - j$, and for the BGM_D model we have the changepoint location intervals $b_1 \in [j, j + 1]$, $j = 2, 3, \dots, 25$. Figures 10(c) and 10(d) show the resulting prior probability ratios in dependence on the changepoint location j . The figure reveals contrary trends for the BGM and the BGM_D model. Figure 10(c) shows that the prior ratio of the BGM_D model peaks in the middle, that is for equally long time series segments, whereas it decreases monotonically with increasing asymmetry of the segment lengths. The BGM model (Figure 10(d)) exhibits the converse behaviour: the

more disproportionate the segment lengths, the higher the prior ratio R . This behaviour becomes even more obvious when varying the time series length m and the breakpoint location j simultaneously. Figure 11 shows the logarithmic prior probabilities in dependence on both: the time series length m and the segment length proportions. For the BGM model (Figure 11(a)) the penalty term $P(V|K)$ increases drastically with the sample size m if both segments are of a certain length. The higher the sample size the stronger the penalty for symmetric segment lengths. Only if j is either very low or very high, that is if the segment lengths are strongly disproportionate, does the size of the prior probability ratio not change drastically with the sample

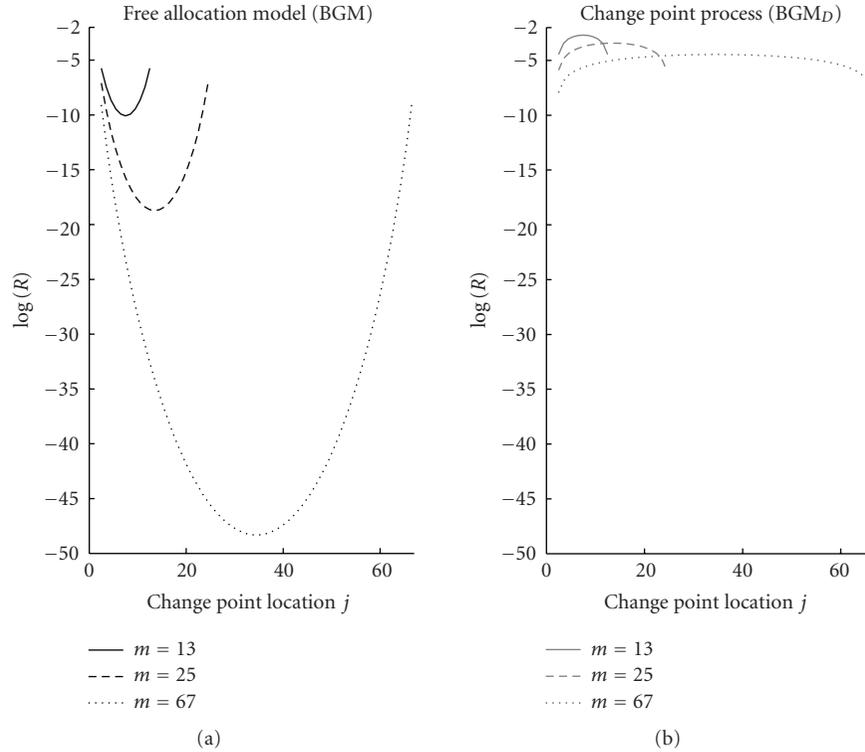


FIGURE 11: Prior probability ratios in dependence on both: the time series length m and the segment lengths proportions. Graphical representation of the prior probability ratio R between (i) the heterogeneous state consisting of two segments t_2, \dots, t_j and t_{j+1}, \dots, t_m and (ii) the homogeneous state consisting of one single segment t_2, \dots, t_m only. The prior probability ratio was defined in (19). For three different lengths of the time series $m = 13, 25, 67$ the logarithmic prior probability ratios (vertical axis) between these two states are plotted in dependence on the changepoint location (horizontal axis). (a) BGM model. (b) BGM_D model.

size m . Figure 11(b) shows that the prior probability ratio of the proposed BGM_D model is less sensitive to both: the sample size m and the segment length proportions. We can conclude that it is the equability of the prior ratio of the proposed BGM_D that renders it superior when modelling nonstationary behaviour in long time series. That is, the penalty for dividing a time series into segments does not change drastically with the length of the time series m , and symmetric segment lengths are supported by the prior distribution $P(V|K)$. On the contrary, the penalty term of the BGM model for dividing a time series into segments increases substantially with the length of the time series: the longer the time series, the lower the prior probability $P(V|K)$. Only if the segment lengths are strongly asymmetric such that one segment is very long and the other very short, is the prior probability $P(V|K)$ of a comparable size for different time series lengths m . This tendency provides a possible explanation for the failure of the BGM model on the *Drosophila melanogaster* gene expression time series. The heat map in Figure 7 shows that the proposed BGM_D model divides the time series into segments that are consistent with the morphogenesis of *Drosophila melanogaster*. The BGM model also tends to detect the correct segment boundaries. However, it then erroneously infers short segments consisting of a only few time points around these boundaries. This

pattern is *not* consistent with the morphogenetical findings. The segmentation inferred with the BGM model thus suffers from artefacts that are an immediate consequence of what has been discussed above. Namely, that the BGM model penalises symmetric partitions much more strongly than asymmetric partitions and thereby encourages the formation of segments that consist of a single or only a few time points. Note that the BGM model becomes more susceptible to these artefacts as the length of the time series increases, as demonstrated in Figure 11(a). This explains why the difference between the BGM and the BGM_D model is less pronounced for the Arabidopsis and macrophage expression time series (Figures 3 and 4), which are considerably shorter (Arabidopsis: 13 time points, macrophages: 25 time points, *Drosophila*: 67 time points). The susceptibility of the BGM model to short-segment artefacts is not completely avoided here either, though, as can be seen from the heat map in Figure 3(b): While BGM_D infers three segments for the macrophages data under condition IFN_γ (see Figure 3(e)), BGM tends to allocate a single data point to a separate segment.

On the synthetic sinusoid data, the proposed BGM_D model yields a higher network reconstruction accuracy than the BGM model, as the latter is more susceptible to inferring spurious self-feedback loops. This tendency can be explained

from the previous mathematical analysis. Both the BGM and the BGM_D model effectively approximate the nonlinear function by a piecewise linear function. A good approximation of the sine wave requires three segments of approximately the same length, corresponding to the ascending, stationary and descending phase. As opposed to BGM_D , the prior inherent to the BGM model heavily penalises against this equal-length segmentation (Figures 10 and 11); see the previous discussion. Now, from (16) it becomes clear that the data are strongly autocorrelated. More precisely, the $Y(t)$'s tend to exhibit a strong autocorrelation by virtue of the autocorrelation of the $X(t)$'s and the influence of $X(t)$ on $Y(t+1)$. Given that the prior implicit in the BGM model impedes the proper piecewise linear model with equal-length segments, the BGM model tends to infer the second-best explanation of the data: explaining the realisation of the $Y(t)$'s via a direct modelling of the autocorrelation between the Y_t 's themselves. This corresponds to (spurious) self-feedback loops.

We note that the novel BGM_D model has been particularly designed for dynamic data with a temporal structure. The BGM model is not restricted to such data, and can equally be applied to both static (steady-state) and dynamic (time series) data. However, the greater flexibility of the BGM model and the intrinsic implications for the effective prior probability on segment lengths and numbers renders its application to time series data suboptimal. This suggests that the proposed BGM_D offers a useful new tool for the analysis of dynamic processes.

7. Conclusions

Two paradigms for relaxing the nonhomogeneity/non-linearity restriction of dynamic Bayesian networks have been proposed in the literature: the changepoint process and the free allocation mixture model. The latter provides the proper approximation of a nonlinear regulation process by a piecewise linear process. The former provides a similar approximation, but under the assumption that the temporal processes are sufficiently smooth, as the assignment of observations to mixture components of the model is done in the temporal domain rather than the domain of regulatory variables. It is obvious that inference in the free allocation model has a considerably higher computational complexity than the changepoint process. However, we have additionally discussed several principled shortcomings that are intrinsic to the methodology per se. We have proposed a new variation of the BGM model [4], which has turned out to be more suitable for the reconstruction of regulatory networks from nonstationary gene expression time series. Like the BGM model the new BGM_D model is based on a mixture model dividing the data points into compartments. The network structures are kept fixed among time series segments, and each segment is modelled separately and independently with the Gaussian BGe model for Bayesian networks. The methodological difference is that the BGM_D model employs a changepoint process to divide time series into segments instead of a free unrestricted allocation of data points. The practical inference follows the Bayesian paradigm and

samples the network, the number of changepoints and the changepoint locations from the posterior distribution with Markov chain Monte Carlo (MCMC).

In a first step, the inference problem was based on synthetic data from small network domains possessing self-feedback loops. Our empirical results show that the proposed BGM_D model suppresses spurious self-feedback loops and yields a higher network reconstruction accuracy than the standard BGe model or the BGM model. We also cross-compared the performance of the models on three real biological systems. On gene expression, time series related to (i) viral challenge of macrophages and (ii) circadian rhythms in *Arabidopsis* heat maps of the connectivity scores revealed that the new BGM_D model infers the biologically expected two-stage structures ((i) dichotomy between the healthy and diseased state of the cell and (ii) the diurnal contrast between light and darkness) more clearly than the BGM model. For assessing the statistical significance of the improvement we focused on predictive probabilities, and the proposed BGM_D model yields consistently higher scores than the BGM model. We extracted a gene regulatory network for the circadian clock-regulated genes in the *Arabidopsis thaliana* domain from the BGM_D inference results, and the reconstructed network shows features that are consistent with the biological literature.

Furthermore, for an independent comparison we combined the free allocation model and the changepoint process model with the discrete multinomial BDe scoring metric for Bayesian networks. Empirical results on binary gene expression time series related to muscle development in *Drosophila melanogaster* were consistent with the results from the first study on continuous data. The (discrete) changepoint process model (BDM_D) infers a time series segmentation that is more consistent with the morphogenetical stages in *Drosophila melanogaster* than the free allocation model (BDM).

We note that the ideal approach—from a biological point of view—for these three applications would be to use a supervised approach, for example, as described in Werhli and Husmeier [27], and to exploit the biological knowledge we have about the experimental conditions (e.g., the morphogenetical stages of *Drosophila*) for data segmentation before inference. However, we elected to use these data as a test case for evaluating the efficiency of unsupervised learning for the proposed changepoint-process DBN model.

Finally, we cross-compared both models from a theoretical point of view. The theoretical study was based on a comparison of (i) the a priori imposed temporal connectivity structures and (ii) the prior probability ratios between the heterogeneous and the homogeneous states of the models. These results are consistent with our empirical findings, and lead to a deeper insight into the intrinsic difference between the two models.

Acknowledgments

M. Grzegorzczk is supported by the Graduate School “Statistische Modellbildung” of the Department of Statistics,

TU Dortmund University. D. Husmeier is supported by the Scottish Government Rural and Environment Research and Analysis Directorate (RERAD).

References

- [1] C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead, "A primer on learning in Bayesian networks for computational biology," *PLoS computational biology*, vol. 3, no. 8, article e129, 2007.
- [2] N. Friedman and D. Koller, "Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks," *Machine Learning*, vol. 50, no. 1-2, pp. 95–125, 2003.
- [3] M. Grzegorzczak and D. Husmeier, "Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move," *Machine Learning*, vol. 71, no. 2-3, pp. 265–305, 2008.
- [4] M. Grzegorzczak, D. Husmeier, K. D. Edwards, P. Ghazal, and A. J. Millar, "Modelling non-stationary gene regulatory processes with a non-homogeneous Bayesian network and the allocation sampler," *Bioinformatics*, vol. 24, no. 18, pp. 2071–2078, 2008.
- [5] A. Nobile and A. T. Fearnside, "Bayesian finite mixtures with an unknown number of components: the allocation sampler," *Statistics and Computing*, vol. 17, no. 2, pp. 147–162, 2007.
- [6] M. A. Suchard, R. E. Weiss, K. S. Dormann, and J. S. Sinsheimer, "Inferring spatial phylogenetic variation along nucleotide sequences: a multiple changepoint model," *Journal of the American Statistical Association*, vol. 98, no. 462, pp. 427–437, 2003.
- [7] M. Grzegorzczak and D. Husmeier, "Modelling non-stationary gene regulatory processes with a non-homogeneous dynamic Bayesian network and the change point process," in *Proceedings of the 6th International Workshop on Computational Systems Biology*, T. Manninen, C. Wiuf, H. Lähdesmäki, et al., Eds., vol. 48 of *TICSP Series*, pp. 51–54, Tampere, Finland, 2009.
- [8] S. Imoto, S. Kim, T. Goto et al., "Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network," *Journal of Bioinformatics and Computational Biology*, vol. 1, no. 2, pp. 231–252, 2003.
- [9] Y. Ko, C. Zhai, and S. L. Rodriguez-Zas, "Inference of gene pathways using Gaussian mixture models," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM '07)*, pp. 362–367, Fremont, Calif, USA, November 2007.
- [10] Y. Ko, C. X. Zhai, and S. Rodriguez-Zas, "Inference of gene pathways using mixture Bayesian networks," *BMC Systems Biology*, vol. 3, article 54, 2009.
- [11] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [12] J. W. Robinson and A. J. Hartemink, "Non-stationary dynamic Bayesian networks," in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds., pp. 1369–1376, 2009.
- [13] S. Lèbre, *Analyse de processus stochastiques pour la génomique : étude du modèle MTD et inférence de réseaux bayésiens dynamiques*, Ph.D. thesis, Université d'Evry-Val-d'Essonne, 2008.
- [14] D. Geiger and D. Heckerman, "Learning Gaussian networks," in *Proceedings of the 10th Annual Conference on Uncertainty in Artificial Intelligence (UAI '94)*, R. López de Mántaras and D. Poole, Eds., Morgan Kaufmann, Seattle, Wash, USA, July 1994.
- [15] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, 1992.
- [16] D. Heckerman and D. Geiger, "Learning Bayesian networks: a unification for discrete and Gaussian domains," in *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence (UAI '95)*, P. Besnard and S. Hanks, Eds., Morgan Kaufmann, Montreal, Canada, August 1995.
- [17] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.
- [18] D. Madigan and J. York, "Bayesian graphical models for discrete data," *International Statistical Review*, vol. 63, pp. 215–232, 1995.
- [19] K. Honda, A. Takaoka, and T. Taniguchi, "Type I interferon gene induction by the interferon regulatory factor family of transcription factors," *Immunity*, vol. 25, no. 3, pp. 349–360, 2006.
- [20] P. A. Salomé and C. R. McClung, "The Arabidopsis thaliana clock," *Journal of Biological Rhythms*, vol. 19, no. 5, pp. 425–435, 2004.
- [21] P. Más, "Circadian clock function in Arabidopsis thaliana: time beyond transcription," *Trends in Cell Biology*, vol. 18, no. 6, pp. 273–281, 2008.
- [22] M. N. Arbeitman, E. E. M. Furlong, F. Imam et al., "Gene expression during the life cycle of *Drosophila melanogaster*," *Science*, vol. 297, no. 5590, pp. 2270–2275, 2002.
- [23] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, pp. 457–472, 1992.
- [24] D. Husmeier, "Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks," *Bioinformatics*, vol. 19, no. 17, pp. 2271–2282, 2003.
- [25] C. R. McClung, "Plant circadian rhythms," *Plant Cell*, vol. 18, no. 4, pp. 792–803, 2006.
- [26] J. C. Locke, M. M. Southern, L. Kozma-Bognár et al., "Extension of a genetic network model by iterative experimentation and mathematical analysis," *Molecular Systems Biology*, vol. 1, article 13, 2005.
- [27] A. V. Werhli and D. Husmeier, "Gene regulatory network reconstruction by bayesian integration of prior knowledge and/or different experimental conditions," *Journal of Bioinformatics and Computational Biology*, vol. 6, no. 3, pp. 543–572, 2008.